PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	IN SITU: Evaluation of the feasibility and impacts of in situ simulation in emergency medicine: a mixed method study protocol
AUTHORS	Truchot, Jennifer; Boucher, Valérie; Raymond-Dufresne, Éliane; Malo, Christian; Brassard, Éric; Marcotte, Jean; Martel, Guillaume; Côté, Geneviève; Garneau, Christian; Bouchard, Gino; Emond, Marcel

VERSION 1 – REVIEW

REVIEWER	Peter Musaeus Aarhus University, Centre for Health Sciences Education
REVIEW RETURNED	07-Sep-2020

GENERAL COMMENTS	Thank you for undertaking a difficult and interesting study. My suggestions are two-fold. First, do use all literature on the topic. Second, reconsider your outcome measures, only patient safety speaks to the system and the more intrapsychological (stress etc.) are perhaps not that relevant. We want to know how in-situ can work in practice, not that participants (who are sturdy and selected through specialty for psychological robustness) are not so stressed out by the unnanounced condition, only a little stressed etc. This in my opinion is not so interesting to warrant a big study. My colleagues and I have published a study on in-situ simulation in EM that you do not cite and you do not cite the literature that we cite: https://pubmed.ncbi.nlm.nih.gov/29450005/ (Designing in situ simulation in the emergency department: evaluating safety attitudes amongst physicians and nurses C Paltved, AT Bjerregaard, K Krogh, JJ Pedersen… - Advances in Simulation, 2017) The authors claim that theirs is the first scientific study to assess the feasibility and impact of implementing In-Situ simulation. This must mean one of three things. First, they deem our published study unscientific or second option, they are unaware of it. Alternatively, they judge our study irrelevant to feasibility. I read their protocol as if they are unaware of our study and in fact some of the literature pertaining to in-situ simulation. More work should go into the design and outcome measures. Are you really measuring feasibility just by measuring psychological (even intra-psychological) variables like stress (which types of stress?), self-confidence and psychosocial risk? In addition, these psychological variables are likely to be co-correlated, what do you do to control for this problem? Where are your measures of feasibility? You need to review the usability and IO psychological literature better. The qualitative part: You state that you will use thematic content

	analysis. The authors mention three levels of coding (open, axial, selective). This sound more like grounded theory than thematic analysis or content analysis. So are they making a new synthesis here? The authors jump from coding, categorizing to theorizing "Then the theory will be created via selective coding". Sorry, no! No theory will necessarily stem from any coding (selective or otherwise). Are you interesting in building a theory? A theory about what? Health Professionals' acceptance of In-situ simulation? What would it mean to build a fully-fledged (social) theory about this phenomenon? I do not follow the steps in your plan in the qualitative empirical part. Simulations: "training will take place in the trauma...area....riskos of mixing up real and false medication. This system will not only enhance realism [because you train participants in the trauma bay?] but also maximize safety [meaning what exactly?] reference 23 [says what about maximizing safety in your study?] POPULATION: You will form training teams of seven health professionals. Are they ad hoc groups? Why groups of seven? What about normal procedures and collaboration standards and procedures? Statistics...you make multiple comparisons, but have no plan for correcting for this problem. Content: You use the word pragmatic on several occasions "to offer a pragmatic and useful training". How do you distinguish between pragmatic and useful? The very first sentence: "Simulation is an innovative teaching tool used for acquisition and training" (of skills). Innovative how? It is not new, so what do you mean? Acquisition AND training? You train so your participants acquire skills (and attitudes as in our study). So what do you mean? "EM is a complex specialty that requires multiple technical and non-technical clinical skills and knowledge." Moreover, attitudes! What do you mean by multiple? Several domains? Several kinds? Your usage of psychological vocabulary could be more accurate. "It also has become increasingly difficult to recruit participants for specific and repeated training". What do you mean by specific? Is it easier to recruit participants for general training? Repeated training (of say 1 hour) obviously is more taxing than training people twice. However, is repeated training half an hour twice more taxing (to whom? the system? the department? the individual? the trainer?) than training once for hone hour? This point seems under-theorized, a bit too every day. The same goes for the fascination with the term realism "also called fidelity". Fidelity is not a one to one mapping of environmental cues and hence the discussion from the perspective of educational science becomes simplistic. Not least because it uncritically makes you assume that fidelity stems from training at the department. Having a good scenario is arguably more important than where it is trained! You reiterate several times that the first leg of the study population will receive in-situ unannounced. Completely unannounced or will they know that this study is going on, i.e. is likely to hit them? Therefore, it is not so unannounced or what? You seem to focus your study on the issue of announced versus unannounced (do participants feel more stressed etc.). Does this really matter to in-situ training or would it be more important to know what worked in in-situ, which factors were active in creating a learning effect in the long run. Please consider having effect measures that speak to learning transfer.

REVIEWER	Vsevolod (Sev) Perelman University Of Toronto
REVIEW RETURNED	19-Sep-2020

GENERAL COMMENTS	The proposed study aims to investigate an important area of quality assurance, continuing professional development and training for emergency medicine practitioners and teams. The objectives of the study include assessing whether in situ simulation is feasible, safe and beneficial for both staff and patients in an academic medium volume emergency department (ED) . The study will include both qualitative and quantitative evaluations of several components and outcomes of in situ simulation interventions. The study has been approved by an institutional ethics review board. The introduction highlights the compelling reasons to conduct in situ simulations involving actual clinical teams at their actual working environment which are supported by the existing literature. The stated hypothesis is that in situ simulations in general in the emergency department are feasible, safe, and result in benefits to clinicians and patients. Additionally, the authors are planning to study both announced and unannounced versions of the interventions. It is my understanding that the following is the list of research questions the study poses to answer: 1. Are the announced in situ simulations in the ED acceptable by the staff? 2. Are the unannounced in situ simulations in the ED acceptable by the staff? 3. Are the announced in situ simulations in the ED feasible? 4. Are the unannounced in situ simulations in the ED feasible? 5. Are the announced in situ simulations in the ED stressful for participants? 6. Are the unannounced in situ simulations in the ED stressful for participants? 7. Are the announced in situ simulations in the ED satisfactory for participants? 8. Are the unannounced in situ simulations in the ED satisfactory for participants? 9. What is the psychological risk to the participants according to their exposure to in situ simulations? 10. What are the technical and non-technical skills of each participant during in situ simulations? 11. What is the number of latent safety threats identified during training? 12. Are the announced in situ simulations in the ED safe for patients? 13. Are the unannounced in situ simulations in the ED safe for patients? I appreciate how challenging it is to do a study, especially utilising in situ simulations in a busy emergency department. The logistics of organizing such a study make it very tempting to use all sorts of data that can be potentially generated by such an undertaking for research purposes.

	While I appreciate that the study can potentially generate a lot of data, my concern is that the authors are attempting to study too many things at the same time. The challenge is that every research question they are raising likely demands different methodology, approach, sample size calculation and statistics. In phase one, the authors are planning to assess “acceptability, safety and prove the validity of the concept”. I anticipate that the authors are experienced and well credentialed simulation educators, however their formal simulation training and certifications are not explicitly stated. I am concerned about some of the elements of simulation design. For example, some of the drugs used will be real, some – simulated. The use of real drugs needs to be justified: There are many very realistic products that are clearly labelled and can be used safely during in situ simulations that will avoid wasting real drugs and will prevent confusion between the real and the simulated drugs (e.g. opioids and blood products). Using simulated drugs (e.g. saline instead of the blood will require having a collaborator/confederate amidst the simulation team. How is that addressed? How is the massive transfusion protocol simulated when blood is simulated by saline (instead of a coloured water), etc.? I am concerned that the scenarios are quite complex and heterogeneous requiring different time commitments. The authors state that simulation will take 15 min. However, a realistic resuscitation of thoracic trauma with the time allocated for interventions and imaging is very different from cardiac arrest. The former will take more than 15 min (30 min to an hour), while the latter 15-20 min. The first one may be overwhelming to less experienced personnel, while the latter is much more habitual, considering the annual BLS and ACLS mandatory training. Addition of the massive transfusion protocol and traumatic brain injury cases adds to the heterogeneity of the experiences for participants in different scenarios. Considering that only 8 sessions will be conducted in each arm, I would advocate to choosing only 2 scenarios of similar complexity. It is stated that simulations will be followed by a 15 min debriefing. In my experience with in situ simulations, to have a meaningful debriefing of 7 people in 15 minutes is an unattainable goal. This time allocation will not allow for a meaningful reflection as each member of 7 people team will only have 2 minutes to speak. Perhaps a plus-delta or a combination of an immediate short debrief and delayed detailed debriefing should be used? I am not clear on the statement that the experience will offer “pragmatic and useful training format”. I am not sure what criteria for pragmatism and usefulness authors are using to make such a statement. The authors are not explicitly describing what in their opinion the feasibility criteria are, which should be defined a priory to test the hypothesis. Does feasibility assessment include the cost of each session? What is the anticipated number of sessions per unit of time (e.g. per month or per year) that is required to develop and maintain certain technical and/or non-technical skills and competencies in the individuals (e.g. team leaders) or teams at large (e.g. for high stake

	resuscitations)? Does feasibility assessment address the cost of additional disposables, cleaning time in the time of COVID, risks of cross contamination and infections, PPEs, etc. What are the a priory criteria of “non-feasibility”? It appears that the authors will gather the number of cancelled sessions and the reasons for their cancellations. However, they will conduct the simulations in the “light hours” of the ED, with the ability to withhold the simulation or stop it on demand. Each simulation will involve seven participants. I would be interested to know how many RNs, RTs, and MDs are usually present in the department during the time of anticipated simulations. Is there any planned or established redundancy of the medical personnel (e.g. on-call personnel)? It also seems that the announced sessions will take place after the end of the shifts for at least some participants. That limits the face validity of the entire experience in terms its effect on many variables the study attempts to investigate. The criteria for aborting the simulations are not explicitly established. Lack of clarity with respect to the above descriptors will limit the external validity and reproducibility of their results. The acceptance of the intervention by the participants will be evaluated by the semi-structured interviews. It is stated to be the primary objective, while the psychological stress, and satisfaction assessment fall into the secondary objectives. I submit that all of them are interlaced. I suspect that overall acceptance by participants will be somewhat dependant on the level of stress and satisfaction. The latter two are dependent on the prior simulation experiences and the quality of the simulations and debriefing. However, there is no description of the attempt to analyze the previous simulation experience of the participants or standardize it with some training or orientation prior to commencing the study. If the authors are not planning to do it, it is possible that different teams will have different simulation experience and that alone may confound the results given a very small number of simulations planned. The prior experience and familiarity with the equipment may be very important, especially in the scenario where they will be using a “prototype for thoracic interventions”. Will all the personnel be familiarized with this device? One way of mitigating the impact of prior experiences is to conduct simulation sessions for all the participants in the simulation centre/lab before the in situ study takes place, but it has its own limitations and feasibility issues. I am not sure at which point the perceived acceptability of the simulations will be analyzed. Will it be done before and after the simulations take place or only after the in situ experience? What will be the acceptable delay between the experience and questionnaire or small group discussions and how will the re-call bias will be mitigated? The attempts to analyze the announced vs un-announced simulations adds even more complexity to the study. It seems that the announced sessions will take place after the shift, requiring the participants to stay after hours. Will the nurses require to be paid as it would be the case in Ontario according to the nursing union rules or not? How will it affect the feasibility, acceptability and satisfaction? The authors did not seem to be utilising pre-post design of the questionnaire, where each participant can be used as a control (pre-sim) for the analysis of the impact (post-sim) self-confidence, self-

	perceived stress, etc. There are many examples when uniformly post simulation, participants recorded high appreciation of the experience and increased self-confidence ("Bobby likes it" phenomenon). It is not clear if the authors are planning post-simulation debriefing and whether the questionnaires will be administered prior to the debriefing or after (which in itself is a challenging decision as debriefing itself is a very valuable element of simulation experience. Additionally, the type and quality of the debriefing process biases the questionnaire responders because: 1. it changes the perception of one's own experiences (e.g. if everybody says it was stressful, who am I to disagree or vice versa) and 2. if the person initially is not aware of his/her/their mistake – the perception of the simulation will change after the debriefing when the errors comes to light. The safety of simulation is a very broad concept as well. It seems that the authors want to evaluate safety to the patients as well as safety to the participants. With regards to evaluating the impact of in situ simulation on patient safety, I do not see that the study has a design, well defined criteria and power for doing so. From the statistical analysis point of view, if only 8 sessions are conducted, the relative impact of one cancellation is big. If cancellations are the dependable variables, one need to analyze the number needed to study that parameter to calculate the power. Similarly, if the anticipated duration of simulation is less than an hour, it is hard to imagine that it will affect the number of patients left without being seen. What is there baseline? Is it a common problem? In terms of the effect on wait time, the authors are planning to conduct the simulations during "light hours". I wonder what the anticipated effect is? How many simulations does one need to see the difference? Does one even need to conduct simulations to assess it? One can ask 7 clinicialns to just sit down for an hour and study the effect of their "absence" from the department. The methodology of defining latent safety threats and analyzing non-technical and technical skills is vague. Will the sessions be video recorded? Who will the rates be? Are they going to be blinded to announced vs unannounced scenarios? In summary, I think the topic of in situ simulation in the Emergency Department is very important. Many questions that authors are attempting to address are very relevant. I think the study has to be revised and simplified, metrics, criteria, outcome measures and methods need to be better defined.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Thank you for undertaking a difficult and interesting study. My suggestions are two-fold. First, do use all literature on the topic. Second, reconsider your outcome measures, only patient safety speaks to the system and the more intrapsychological (stress etc.) are perhaps not	Thank you so much for your comments. We are also conducting a systematic review on the use of in situ simulation in emergency medicine and therefore are familiarized with the vast number of research articles addressing this issue (more than 100). We indeed were unable

that relevant. We want to know how in-situ can work in practice, not that participants (who are sturdy and selected through specialty for psychological robustness) are not so stressed out by the unnanounced condition, only a little stressed etc. This in my opinion is not so interesting to warrant a big study.	to cite every one of these studies in the references list of our protocol. Following your comment, we have added some references we thought were important to clarify our methodology. As for the main objective of our work, safety is assessed through quantitative assessment of various parameters (the ED median wait time 6 hrs before and 6 hrs post ISS (Stretcher and ambulatory care separately) and 2) the number of patients who left without being seen or against medical advice 6h post ISS.) During the semi directed interviews, we also address the issue of safety for patients and for participants. Therefore, our feasibility methodology includes a safety analysis. We have modified and revised the manuscript extensively, to gain clarity and highlight the importance of "safety" assessment in our study.
My colleagues and I have published a study on in-situ simulation in EM that you do not cite and you do not cite the literature that we cite: https://pubmed.ncbi.nlm.nih.gov/29450005/ (Designing in situ simulation in the emergency department: evaluating safety attitudes amongst physicians and nurses C Paltved, AT Bjerregaard, K Krogh, JJ Pedersen… - Advances in Simulation, 2017) The authors claim that theirs is the first scientific study to assess the feasibility and impact of implementing In-Situ simulation. This must mean one of three things. First, they deem our published study unscientific or second option, they are unaware of it. Alternatively, they judge our study irrelevant to feasibility. I read their protocol as if they are unaware of our study and in fact some of the literature pertaining to in-situ simulation.	Thank you for sharing your work. We have read your study with much attention. Your approach to this question has helped us rethink and clarify our research methodology. We have added your study to the references list of our protocol.
More work should go into the design and outcome measures. Are you really measuring feasibility just by measuring psychological (even intra-psychological) variables like stress (which types of stress?), self-confidence and psychosocial risk? In	We decided to measure feasibility according to two modalities: - Qualitative assessment via semi-structured

addition, these psychological variables are likely to be co-correlated, what do you do to control for this problem?	individual interviews - Measuring the number of cancelled sessions and the number of LST identified through both modalities (announced/unannounced) These numbers should be equivalent; otherwise this will favour the feasibility of one format over the other. To clarify these elements, the text has been revised
Where are your measures of feasibility? You need to review the usability and IO psychological literature better.	We have clarified our methodology as per your comment. The measures of feasibility are cited in the previous response.
The qualitative part: You state that you will use thematic content analysis. The authors mention three levels of coding (open, axial, selective). This sound more like grounded theory than thematic analysis or content analysis. So are they making a new synthesis here? The authors jump from coding, categorizing to theorizing "Then the theory will be created via selective coding". Sorry, no! No theory will necessarily stem from any coding (selective or otherwise). Are you interesting in building a theory? A theory about what? Health Professionals' acceptance of In-situ simulation? What would it mean to build a fully-fledged (social) theory about this phenomenon? I do not follow the steps in your plan in the qualitative empirical part.	We thank you for these remarks and have corrected accordingly this paragraph after discussing this with the research team (see thematic content analyses p.16)
Simulations: "training will take place in the trauma...area....riskos of mixing up real and false medication. This system will not only enhance realism [because you train participants in the trauma bay?] but also maximize safety [meaning what exactly?] reference 23 [says what about maximizing safety in your study?] POPULATION: You will form training teams of seven health professionals. Are they ad hoc groups? Why groups of seven? What about normal procedures and collaboration standards and procedures?	We have modified the text to clarify these issues. In order to ensure patient safety and limit the risks of disrupting patient care in the ED during ISS, the simulation experts have designed specific “go/no go” criteria. The go and no go criteria have been inspired by the existing literature on ISS in clinical settings such as the ED and we have adapted this list to some organisational specificities of our ED set up bajaj go nogo, reamer et al safety first). These no go criteria include: heavy clinical load, work flow understaffing, bed availability on wards and equipment needs (for example, unavailability of the fast flow fluid warmer). . also if a real trauma patient is expected or ongoing, the ISS will be

	cancelled, and the simulation team will leave the trauma room in the same state as it was upon arrival. This system will not only enhance realism but also insure safety.[23]» Reference 23 is an editorial, helping those organising in situ simulation, to respect to ethical and safety issues associated to training in a real clinical settings. The authors even wrote the ten commandments of safe simulation, advising educators to abide them. Population: « We chose to include teams of seven participants because it corresponds to the exact number of participants involved in the trauma team when a real critical patient is admitted to the resuscitation/trauma area of our ED. »
Statistics...you make multiple comparisons, but have no plan for correcting for this problem	This information was added to our manuscript (see quantitative analyses section)
Content: You use the word pragmatic on several occasions "to offer a pragmatic and useful training". How do you distinguish between pragmatic and useful?	We have removed this word from the text, because of the ambiguity of this term.
The very first sentence: "Simulation is an innovative teaching tool used for acquisition and training" (of skills). Innovative how? It is not new, so what do you mean? Acquisition AND training? You train so your participants acquire skills (and attitudes as in our study). So what do you mean? "EM is a complex specialty that requires multiple technical and non-technical clinical skills and knowledge." Moreover, attitudes! What do you mean by multiple? Several domains? Several kinds? Your usage of psychological vocabulary could be more accurate.	We have modified the introduction and rewritten these sentences to be more accurate.
"It also has become increasingly difficult to recruit participants for specific and repeated training". What do you mean by specific? Is it easier to recruit participants for general training? Repeated training (of say 1 hour) obviously is more taxing than training people twice. However, is repeated training half an hour twice more taxing (to whom?	We have clarified these elements by rewriting this paragraph according to existing literature.

the system? the department? the individual? the trainer?) than training once for hone hour? This point seems under-theorized, a bit too every day. The same goes for the fascination with the term realism "also called fidelity". Fidelity is not a one to one mapping of environmental cues and hence the discussion from the perspective of educational science becomes simplistic. Not least because it uncritically makes you assume that fidelity stems from training at the department. Having a good scenario is arguably more important than where it is trained!	
You reiterate several times that the first leg of the study population will receive in-situ unannounced. Completely unannounced or will they know that this study is going on, i.e. is likely to hit them? Therefore, it is not so unannounced or what? You seem to focus your study on the issue of announced versus unannounced (do participants feel more stressed etc.). Does this really matter to in-situ training or would it be more important to know what worked in in-situ, which factors were active in creating a learning effect in the long run. Please consider having effect measures that speak to learning transfer.	The simulation will be unannounced (during working hours) or announced (before working hours). Some of the emergency professionals might suspect a study is going on, but they will not know the moment they might be exposed to ISS. This is an important issue for us, because it will help us decide which format is best, not only according to the professionals (which format they prefer), but also for the safety of real ED patients. This is important to justify the long-term application of one or the other, to the hospital board. The publication of this scientific process can also help other educators choose the right format of ISS.

Reviewer 2

It is my understanding that the following is the list of research questions the study poses to answer: 1. Are the announced in situ simulations in the ED acceptable by the staff? 2. Are the unannounced in situ simulations in the ED acceptable by the staff? 3. Are the announced in situ simulations in the ED feasible? 4. Are the unannounced in situ simulations in the ED feasible? 5. Are the announced in situ simulations in the ED stressful for participants? 6. Are the unannounced in situ simulations in the	You understand correctly and indeed we aim at answering questions 1, 2, 3, 4 during phase 1. This will be the main outcome: feasibility of ISS During phase 1, the secondary outcome will include the question 12 and 13 The phase 2 of this study will explore the impact of ISS on the participants' psychosocial risks, that is question 9, a bit rephrased as "can ISS reduce psychosocial risks among emergency professionals "

ED stressful for participants? 7. Are the announced in situ simulations in the ED satisfactory for participants? 8. Are the unannounced in situ simulations in the ED satisfactory for participants? 9. What is the psychological risk to the participants according to their exposure to in situ simulations? 10. What are the technical and non-technical skills of each participant during in situ simulations? 11. What is the number of latent safety threats identified during training? 12. Are the announced in situ simulations in the ED safe for patients? 13. Are the unannounced in situ simulations in the ED safe for patients?	We have removed questions 5 to 8 from our methodology, to reduce the workload and recenter our research question to feasibility, and safety. We have also removed from our method question 10, it perhaps is not that relevant and has already been explored by other authors. Question 11 will be answered through direct observations of each ISS during phase 1 and 2 and will be included in our feasibility analysis, under the postulate that the number of LST should be equivalent under the 2 modalities (announced/unannounced).
I appreciate how challenging it is to do a study, especially utilising in situ simulations in a busy emergency department. The logistics of organizing such a study make it very tempting to use all sorts of data that can be potentially generated by such an undertaking for research purposes. While I appreciate that the study can potentially generate a lot of data, my concern is that the authors are attempting to study too many things at the same time. The challenge is that every research question they are raising likely demands different methodology, approach, sample size calculation and statistics	Thank you very much for this comment. We have decided to limit the data analysis and we have removed some outcomes and questions from the research strategy. In order to gain clarity in the presentation of our research and reduce the workload and the challenges of this ambitious project.
I anticipate that the authors are experienced and well credentialed simulation educators, however their formal simulation training and certifications are not explicitly stated.	Half the research team, and all of which involved in the field work of this research, are experienced and well credentialed simulation educators (JT, CM, EB, ERD) and have undergone simulation training, fellowship and/or a phd in education sciences.
I am concerned about some of the elements of simulation design. For example, some of the drugs used will be real, some – simulated. The use of real drugs needs to be justified: There are many very realistic products that are clearly labelled and can be used safely during in situ simulations that will avoid wasting real drugs and will prevent confusion between the real and the simulated drugs (e.g. opioids and blood products). Using simulated drugs (e.g. saline instead of the blood will require having a	We have clarified this issue (see simulations section).

collaborator/confederate amidst the simulation team. How is that addressed? How is the massive transfusion protocol simulated when blood is simulated by saline (instead of a coloured water), etc.?	
I am concerned that the scenarios are quite complex and heterogeneous requiring different time commitments. The authors state that simulation will take 15 min. However, a realistic resuscitation of thoracic trauma with the time allocated for interventions and imaging is very different from cardiac arrest. The former will take more than 15 min (30 min to an hour), while the latter 15-20 min. The first one may be overwhelming to less experienced personnel, while the latter is much more habitual, considering the annual BLS and ACLS mandatory training. Addition of the massive transfusion protocol and traumatic brain injury cases adds to the heterogeneity of the experiences for participants in different scenarios. Considering that only 8 sessions will be conducted in each arm, I would advocate to choosing only 2 scenarios of similar complexity.	The literature illustrating the positive impact and the validity of short in situ simulation with short debriefings exists and has been added to justify this choice in our design (see simulations section). We have also opted for practical scenario design, adapted to our local training needs, as many trauma patients are managed in our ED every year. One of the objectives of our study is to improve well-being (reduce psychosocial risks, stress and improve self-confidence) We believe a complete, rich training program would insure reaching that goal better than using the same two scenarios.
It is stated that simulations will be followed by a 15 min debriefing. In my experience with in situ simulations, to have a meaningful debriefing of 7 people in 15 minutes is an unattainable goal. This time allocation will not allow for a meaningful reflection as each member of 7 people team will only have 2 minutes to speak. Perhaps a plus-delta or a combination of an immediate short debrief and delayed detailed debriefing should be used?	We justify this choice on the existing literature (see simulation section). The semi directed interviews will also offer a unique opportunity to discuss and debrief the simulation sessions.
I am not clear on the statement that the experience will offer “pragmatic and useful training format”. I am not sure what criteria for pragmatism and usefulness authors are using to make such a statement.	We have removed the word pragmatic
The authors are not explicitly describing what in their opinion the feasibility criteria are, which should be defined a priory to test the hypothesis. Does feasibility assessment include the cost of each session? What is the anticipated number of sessions per unit of time (e.g. per month or per year) that is required to develop and maintain certain technical and/or non-technical skills and competencies in the individuals (e.g. team leaders) or teams at large (e.g. for high stake resuscitations)? Does feasibility assessment address the cost of additional disposables, cleaning	We do not plan to analyze the cost of these sessions (material, equipment etc.), nor to compare these costs to other simulation formats. This could be an interesting approach to feasibility design, however we have not planned on doing this during our research project. Our ‘’a priory’’ criteria of non-feasibility will

time in the time of COVID, risks of cross contamination and infections, PPEs, etc. What are the a priory criteria of “non-feasibility”?	emerge from the semi directed interviews, and the cancelation rate.
It appears that the authors will gather the number of cancelled sessions and the reasons for their cancellations. However, they will conduct the simulations in the “light hours” of the ED, with the ability to withhold the simulation or stop it on demand. Each simulation will involve seven participants. I would be interested to know how many RNs, RTs, and MDs are usually present in the department during the time of anticipated simulations. Is there any planned or established redundancy of the medical personnel (e.g. on-call personnel)? It also seems that the announced sessions will take place after the end of the shifts for at least some participants. That limits the face validity of the entire experience in terms its effect on many variables the study attempts to investigate. The criteria for aborting the simulations are not explicitly established. Lack of clarity with respect to the above descriptors will limit the external validity and reproducibility of their results.	We have rewritten these sentences to enhance clarity (see population). The announced format will take place before the shift (and not after).
The acceptance of the intervention by the participants will be evaluated by the semi-structured interviews. It is stated to be the primary objective, while the psychological stress, and satisfaction assessment fall into the secondary objectives. I submit that all of them are interlaced. I suspect that overall acceptance by participants will be somewhat dependant on the level of stress and satisfaction. The latter two are dependent on the prior simulation experiences and the quality of the simulations and debriefing.	The assessment of feasibility will take place during phase 1, indeed during the semi directed interviews. During phase 1, we compare two formats announced and unannounced. Whereas the assessment of psychosocial risks will occur during phase 2. We will compare the benefit on psychosocial risks by comparing the group of professionals exposed to ISS to those that have not been exposed to ISS. These endpoints may be interlaced but they will be assessed at a different time point, using different tools and a different population. For each of these evaluations, the notion of prior simulation experiences will be collected, and the results analysed accordingly, with proper adjustments if needed.
However, there is no description of the attempt to analyze the previous simulation experience of the participants or standardize it with some training or orientation prior to commencing the study. If the authors are not planning to do it, it is possible that different teams will have different simulation experience and that alone may confound the results given a very small number of simulations planned.	The previous experience with simulation will be collected during the semi directed interviews and also with the phase 2 questionnaires. This data will be used and analysed with proper adjustments. We added some information regarding the

The prior experience and familiarity with the equipment may be very important, especially in the scenario where they will be using a “prototype for thoracic interventions”. Will all the personnel be familiarized with this device? One way of mitigating the impact of prior experiences is to conduct simulation sessions for all the participants in the simulation centre/lab before the in situ study takes place, but it has its own limitations and feasibility issues.	important question of familiarity with the equipment (see simulation section p. 11).
I am not sure at which point the perceived acceptability of the simulations will be analyzed. Will it be done before and after the simulations take place or only after the in situ experience? What will be the acceptable delay between the experience and questionnaire or small group discussions and how will the re-call bias will be mitigated? The attempts to analyze the announced vs un-announced simulations adds even more complexity to the study. It seems that the announced sessions will take place after the shift, requiring the participants to stay after hours. Will the nurses require to be paid as it would be the case in Ontario according to the nursing union rules or not? How will it affect the feasibility, acceptability and satisfaction?	To answer this query, we have clarified our methodology. We hope that the explanation we provided in the ‘’procedure’’ section appropriately answers this reviewer’s comment.
The authors did not seem to be utilising pre-post design of the questionnaire, where each participant can be used as a control (pre-sim) for the analysis of the impact (post-sim) self-confidence, self-perceived stress, etc. There are many examples when uniformly post simulation, participants recorded high appreciation of the experience and increased self-confidence (“Bobby likes it” phenomenon). It is not clear if the authors are planning post-simulation debriefing and whether the questionnaires will be administered prior to the debriefing or after (which in itself is a challenging decision as debriefing itself is a very valuable element of simulation experience. Additionally, the type and quality of the debriefing process biases the questionnaire responders because: 1. it changes the perception of one’s own experiences (e.g. if everybody says it was stressful, who am I to disagree or vice versa) and 2. if the person initially is not aware of his/her/their mistake – the perception of the simulation will change after the debriefing when the errors comes to light.	Our previous version of the protocol may not have properly illustrated that we aim to compare two population for each phase of the study. We believe this to have a stronger validity than a before after design for the purposes of our study. The intergroup comparison will be the following: Phase 1: comparison between announced and unannounced for feasibility and safety. Phase 2: comparison between exposed to ISS or non-exposed to ISS for the psychosocial risks, the self-confidence and stress questionnaires.

	The questionnaires will be filled out after the simulation (which includes debriefing, of course, otherwise, the endpoints would be quite different).
The safety of simulation is a very broad concept as well. It seems that the authors want to evaluate safety to the patients as well as safety to the participants.	Yes we do. Thanks to your comment, we have added some information in the ‘’safety’’ section (p.13)
With regards to evaluating the impact of in situ simulation on patient safety, I do not see that the study has a design, well defined criteria and power for doing so. From the statistical analysis point of view, if only 8 sessions are conducted, the relative impact of one cancellation is big. If cancellations are the dependable variables, one need to analyze the number needed to study that parameter to calculate the power. Similarly, if the anticipated duration of simulation is less than an hour, it is hard to imagine that it will affect the number of patients left without being seen. What is there baseline? Is it a common problem? In terms of the effect on wait time, the authors are planning to conduct the simulations during “light hours”. I wonder what the anticipated effect is? How many simulations does one need to see the difference? Does one even need to conduct simulations to assess it? One can ask 7 clinicialns to just sit down for an hour and study the effect of their “absence” from the department.	The statistics paragraph has been revised and rewritten to answer these comments (see quantitative analyses section)
The methodology of defining latent safety threats and analyzing non-technical and technical skills is vague. Will the sessions be video recorded? Who will the rates be? Are they going to be blinded to announced vs unannounced scenarios?	We will use tools to identify LST that have been validated in previous studies. The sessions will not be recorded, and a specific observer will be in charge of identifying LSTs (see p. 12)
In summary, I think the topic of in situ simulation in the Emergency Department is very important. Many questions that authors are attempting to address are very relevant. I think the study has to be revised and simplified, metrics, criteria, outcome measures and methods need to be better defined	Thank you for your expert analysis of our protocol. We believe this will be in line with the readers of BMJ open and the results to our numerous questions will facilitate the implementation of ISS in EM.

VERSION 2 – REVIEW

REVIEWER	Vsevolod Perelman The University of Toronto, Canada
REVIEW RETURNED	18-Dec-2020

GENERAL COMMENTS	IN SITU: Evaluation of the feasibility and impacts of in situ simulation (ISS) in emergency medicine - a mixed-method study protocol The proposed study investigates an important area of quality assurance, continuing professional development and training for emergency medicine practitioners and teams. The study hypothesizes that in situ simulations in an academic high-volume emergency department are feasible, safe, and associated with benefits for both staff and patients. The method includes "simultaneously" (page 5, line 3) • a qualitative assessment of feasibility and acceptability (page 5, line 3-4) • a quantitative assessment of patients' safety and participants' psychosocial risks. The study has been approved by an institutional ethics review board and was supported by the Foundation du CHU de Québec-Université Laval grant number 3967 Two distinct phases are planned (page 8) • Phase 1: Objective: To assess and compare the feasibility and safety of announced and unannounced in-situ simulation in the emergency department. • Phase 2: Objective: To assess whether ISS improves participants' psychosocial impact (stress reduction, satisfaction improvement) compared to no exposure. It is not clear what "no exposure" means: during real resuscitations without ISS experience? • An unrelated objective, so-called, secondary objective: To compare the number of latent safety threats (LST) identified during unannounced ISS vs. announced ISS. 1) Phase 1 ISS Feasibility Safety Announced Unannounced 2) Phase 2 Stress reduction Satisfaction Improvement Exposed to ISS Not exposed to ISS 3) Secondary Objective ISS Latent safety threats Satisfaction Improvement Announced Unannounced The stated outcomes are worded slightly differently (p 13, lines 11-27) Primary outcomes: Phase 1: Proportion of successful ISS and qualitative exploration of feasibility among the two groups: announced and unannounced Phase 2: Psychosocial risks levels among the two groups: ISS and no ISS (intervention and control) Secondary outcomes: Phase 1: • Quantitative patient safety parameters (wait times, adverse events, departures without being seen) • Number of LSTs among the two groups: announced and unannounced

	Phase 2: • Self-confidence levels among the two groups: ISS and no ISS • Stress levels among the two groups I suggest the authors use the same terminology consistently: e.g. safety or psychological risk? Patient safety or participants' safety? Self-confidence was not listed in the study objectives. Introduction The introduction has been revised and reads well, highlighting the reasons to conduct and further investigate in situ simulations. Methodology: Pre-assigned interdisciplinary trauma teams of 7 participants will be formed from "selected volunteers." All participants would have had some exposure to simulations and specifically to the manikin being used. In phase 1, all simulations will be announced (page 9, line 24). However, in Fig.1, it seems that some of the simulations will be happening unannounced. Will they be assessed for Safety and Feasibility? Or the figure does not accurately depict the process? On page 13, lines 30-32, the authors specify that "during Phase 1, a total of 16 sessions will be required (8 announced, eight unannounced) with a total of 112 participants". That inconsistency should be clarified. It seems that during phase 2, only unannounced ISS will take place (p 12, line 27) and further (page 14, lines 4-7): "during Phase 2, a total of 10 unannounced ISS will be required to compare the group of participants exposed to ISS (n=70) to the professionals not exposed to ISS (n=70). During phase 2, "participation will be random as the sessions will be unannounced" (p.9, line 25). This needs to be clarified: Does it mean that cases will be randomized, or the groups of the participants will be randomized within the groups? Or that participants will not be "volunteers" any longer? Similar to the previous comment, figures need to be clarified: What does "Simulation (-)" mean in Fig. 2? The rationale for using real medications has been provided: to better identify safety threats. Specific "go/no go "criteria have been defined (perhaps, the requirements need to be listed in the appendix, as those criteria will affect the interference and, therefore, the outcome of the study). Scenarios are traumatic brain injury, penetrating thoracic trauma, massive transfusion protocol activation and cardiac arrest have been tested and addressed "specific local teaching needs" (p10, line 2). Does it mean that some assessment of the performance has already taken place, and "specific teaching needs" have been identified to address deficiencies in performance? Will some of the teaching take place prior to the study to mitigate it? Does it mean that the study already has an inherent bias? I mean that if there is a local need for teaching, it may mean that teams are not performing well. That will

	translate into increase stress and anxiety during in situ simulation because of the nature of the scenario. But authors propose to study, amongst other things, stress and anxiety and the psychological comfort and compare it to "no-exposure" in phase 2, which may not be encountering the same cases. If the goal is to investigate generic in-situ practices, should not the authors study cases that do not "fulfill specific teaching needs" and instead use the "typical" cases that the teams should be familiar with already? The length of simulation/debriefing is adjusted, explained in more detail and the rationale provided. A Bowen-Kreuter framework will be utilized for feasibility assessment in the areas of acceptability, implementation and practicality and is well defined At the end of phase 1, acceptability and practicality will be assessed during semi-structured interviews will be conducted with the themes, designed with the help of a qualitative research specialist. Implementational feasibility will be depicted by the number of cancelled sessions and what sounds like a descriptive analysis of the circumstances and reasons for cancellations. I would add an explanatory comment of the delays, timing of set up and cleaning. The authors proposed to use the number of identified LSTs during the announced and unannounced ISS as part of the feasibility assessment. I submit that it belongs more to the analysis of safety (p 12). Safety assessment seems to be based on the comparison of the pre-ISS and post-ISS waiting time in ED stratified by a triage category and the number of "left without being seen" patients and compared to three-day average parameters as "control" period matched for periods of the 24 hr ED cycle. Page 12: "Dedicated research staff will be present in the ED during the simulations and for up to one hour afterwards in order to record the occurrence of reported patient-related adverse events (accident report)." Will it be done for all the patients in the entire department? Page 12, line 22: "…to record… the impact of the simulations on the working staff (understaffing and work overload)"… How will that be assessed? Phase 2 seems to focus on assessing the impacts of unannounced ISS on health professionals with regards to psychosocial stress, self-confidence and professional wellbeing using validated satisfaction and stress scales. "To demonstrate this benefit, staff members exposed to either unannounced or announced ISS training (intervention group) would be compared to those that were not exposed to ISS (control group) using the same questionnaires." (p13, lines 7-9). This is the first time the researchers introduce the terminology of intervention vs control groups. There are no descriptions of how the intervention group will be selected vs the control. It seems from the description of the phase 1 methodology that the intervention groups will be "selected volunteers." How will the researchers control for a selection bias within the groups? In other words, how will they make sure that people who volunteer for the intervention group have the same psychological responses as the controls?

	Thematic content analysis is not described in detail and needs to be further defined. Specifically: how consistency and cohesion will be promoted? What techniques will be used to address and achieve credibility, transferability of inquiry, and dependability, and establish confirmability? E.g. will there be an independent audit planned? In summary, I think the topic of in situ simulation in the Emergency Department is significant. Many questions that the authors are attempting to address are very relevant. I believe the study protocol still needs to be better clarified and inconsistencies addressed and corrected.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

Two distinct phases are planned (page 8) • Phase 1: Objective: To assess and compare the feasibility and safety of announced and unannounced in-situ simulation in the emergency department. •Phase 2: Objective: To assess whether ISS improves participants' psychosocial impact (stress reduction, satisfaction improvement) compared to no exposure. It is not clear what "no exposure" means: during real resuscitations without ISS experience?	We have clarified the wording and harmonised some terminology throughout the manuscript. We meant by “no exposure” the emergency professionals that will not be exposed to ISS. We have changed the ‘no exposure” to “control” group. This is a large ED, and it is impossible to train with ISS each professional working in the ED, therefore half of the emergency professionals will be trained with ISS and the other half not. The professionals that will not undergo ISS will be called the “control” group.
An unrelated objective, so-called, secondary objective: To compare the number of latent safety threats (LST) identified during unannounced ISS vs. announced ISS. 1) Phase 1 ISS Feasibility Safety Announced Unannounced 2) Phase 2 Stress reduction Satisfaction Improvement Exposed to ISS Not exposed to ISS	We have rephrased and paid attention to the harmonisation of the wording of the outcomes.

3) Secondary Objective ISS Latent safety threats Satisfaction Improvement Announced Unannounced The stated outcomes are worded slightly differently (p 13, lines 11-27)	
I suggest the authors use the same terminology consistently: e.g. safety or psychological risk? Patient safety or participants' safety? Self-confidence was not listed in the study objectives. I suggest the authors use the same terminology consistently: e.g. safety or psychological risk? Patient safety or participants' safety? Self-confidence was not listed in the study objectives.	To clarify these elements, the text has been revised We have clarified our methodology as per your comment and added self-confidence to the psychosocial analysis.
Introduction - The introduction has been revised and reads well, highlighting the reasons to conduct and further investigate in situ simulations. Methodology: Pre-assigned interdisciplinary trauma teams of 7 participants will be formed from "selected volunteers." All participants would have had some exposure to simulations and specifically to the manikin being used.	Thank you very much for these comments
In phase 1, all simulations will be announced (page 9, line 24). However, in Fig.1, it seems that some of the simulations will be happening unannounced. Will they be assessed for Safety and Feasibility? Or the figure does not accurately depict the process? On page 13, lines 30-32, the authors specify that "during Phase 1, a total of 16 sessions will be required (8 announced, eight unannounced) with a total of 112 participants". That inconsistency should be clarified	We thank you for these remarks and have corrected accordingly this paragraph after discussing this with the research team. During phase 1, both formats announced and unannounced will occur and be compared.
It seems that during phase 2, only unannounced ISS will take place (p 12, line 27) and further (page 14, lines 4-7): "during Phase 2, a total of 10 unannounced ISS will be required to compare the	We have modified the text to clarify these issues. It means that during phase 2, only unannounced

group of participants exposed to ISS (n=70) to the professionals not exposed to ISS (n=70). During phase 2, "participation will be random as the sessions will be unannounced" (p.9, line 25). This needs to be clarified: Does it mean that cases will be randomized, or the groups of the participants will be randomized within the groups? Or that participants will not be "volunteers" any longer?	ISS will happen and we have projected that half of the professionals from our ED will be able to participate to these 10 sessions and therefore the other half will not be “exposed to this training” and constitute our natural control group. The participants will indeed no longer be “selected volunteers” .
Similar to the previous comment, figures need to be clarified: What does "Simulation (-)" mean in Fig. 2?	The figure has been modified accordingly; simulation - meaning ‘the control’ group, those that will not participate in the ISS training.
Specific "go/no go "criteria have been defined (perhaps, the requirements need to be listed in the appendix, as those criteria will affect the interference and, therefore, the outcome of the study). Scenarios are traumatic brain injury, penetrating thoracic trauma, massive transfusion protocol activation and cardiac arrest have been tested and addressed "specific local teaching needs" (p10, line 2). Does it mean that some assessment of the performance has already taken place, and "specific teaching needs" have been identified to address deficiencies in performance? Will some of the teaching take place prior to the study to mitigate it? Does it mean that the study already has an inherent bias? I mean that if there is a local need for teaching, it may mean that teams are not performing well. That will translate into increase stress and anxiety during in situ simulation because of the nature of the scenario. But authors propose to study, amongst other things, stress and anxiety and the psychological comfort and compare it to "no-exposure" in phase 2, which may not be encountering the same cases. If the goal is to investigate generic in-situ practices, should not the authors study cases that do not "fulfill specific teaching needs" and instead use the "typical" cases that the teams should be familiar with already?	We have added the list of no go criteria in the appendix As for the scenarios, we believe ISS to be a pragmatic approach to enhance the quality of patient care. Therefore the scenarios must correspond to the type of care delivered in the corresponding structure. Performance assessment was not done ‘’per se’’, however our scenarios correspond to the cases usually seen in this ED. As written on page 9 in the “scenarios design paragraph” « The simulation team tested the scenarios beforehand during dedicated simulation training with **a different population** than the study participants. ». So no, we have avoided this bias. And what we need by teaching needs is to train professionals to clinical cases they actually encounter frequently so this can translate into improved care: « One of the key purposes for designing this study was to fulfil specific local teaching needs. Therefore, our tested training format will be useful to participants and could easily translate into improved patient care. »
Implementational feasibility will be depicted by the number of cancelled sessions and what sounds like a descriptive analysis of the circumstances	Thank you very much for this suggestion, we will add the descriptive analysis of cancellations to

and reasons for cancellations. I would add an explanatory comment of the delays, timing of set up and cleaning. The authors proposed to use the number of identified LSTs during the announced and unannounced ISS as part of the feasibility assessment. I submit that it belongs more to the analysis of safety (p 12).	our data.
Safety assessment seems to be based on the comparison of the pre-ISS and post-ISS waiting time in ED stratified by a triage category and the number of "left without being seen" patients and compared to three-day average parameters as "control" period matched for periods of the 24 hr ED cycle. Page 12: "Dedicated research staff will be present in the ED during the simulations and for up to one hour afterwards in order to record the occurrence of reported patient-related adverse events (accident report)." Will it be done for all the patients in the entire department? - Page 12, line 22: "…to record… the impact of the simulations on the working staff (understaffing and work overload)"… How will that be assessed?	- Yes it will be done for all the patients - The research staff will collect the comments from the remaining staff (those not participating in the simulations), and these comments will be pooled and analysed. See page 12 of the manuscript where we added this clarification.
Phase 2 seems to focus on assessing the impacts of unannounced ISS on health professionals with regards to psychosocial stress, self-confidence and professional wellbeing using validated satisfaction and stress scales. "To demonstrate this benefit, staff members exposed to either unannounced or announced ISS training (intervention group) would be compared to those that were not exposed to ISS (control group) using the same questionnaires." (p13, lines 7-9). This is the first time the researchers introduce the terminology of intervention vs control groups. There are no descriptions of how the intervention group will be selected vs the control. It seems from the description of the phase 1 methodology that the intervention groups will be "selected volunteers." How will the researchers control for a selection bias within the groups? In other words, how will they make sure that people who volunteer for the intervention group have the same psychological responses as the controls?	-Page 9, population we describe that during phase 2 the ISS will be unannounced and therefore the participants randomly chosen (unlike phase 1's selected volunteers) "During Phase 2, participation will be random, as the sessions will be only unannounced. » To avoid any we have replaced intervention group by ISS group throughout the text Afterwards, we will identify and adjust to confounding factors to control inherent bias from our analysis.

Thematic content analysis is not described in detail and needs to be further defined. Specifically: how consistency and cohesion will be promoted? What techniques will be used to address and achieve credibility, transferability of inquiry, and dependability, and establish confirmability? E.g. will there be an independent audit planned?	The qualitative method has been further detailed See ‘’Thematic content analyses’’ sections (p.14-15)

VERSION 3 – REVIEW

REVIEWER	Vsevolod Perelman University of Toronto, Canada
REVIEW RETURNED	01-Feb-2021

GENERAL COMMENTS	Study: IN SITU: Evaluation of the feasibility and impacts of in situ simulation (ISS) in emergency medicine - a mixed-method study protocol I am grateful for the author’s responses to my questions, their clarifications and revisions. I am satisfied with their responses and now have a much clearer understanding of the project. As the general assumption goes, most latent and overt safety threats will occur exactly during the times when system is stretched to the limit. Unfortunately, those will be exactly the times when the authors propose to abort in-situ simulations. I do appreciate that ultimately, patients’ and clinicians’ safety take priority over research and training. Thus, my only remaining suggestion is to clarify or to be more descriptive of what constitutes “No go criteria” that will lead to cancellation of the in -situ simulations? I am following the list provided by the authors in the appendix: • Availability of the environment – I am not sure what the authors mean by that? • Medical Understaffing –How many RNs or MDs or RTs in your department will it take to be absent to declare “understaffing”? I would specify the exact number: e.g. if the department is more than 1 RN short or at least 1 MD did not show up, or an on-call MD needed to be called…etc. • Non-medical Understaffing – Do you mean environmental support workers or administrative staff or scribes? How many of them will it take to declare “understaffing”? Who will decide whether the department I understaffed by non-clinical personnel? Why would it affect a clinical in-situ simulation? Unless you have a strong opinion about that, I would remove that criterion. • Overcrowding - Do the authors have a predetermined wait time in mind? Or the number of people in the waiting room? Or the number of admitted patients in the department? I would specify: e.g. wait time > 2 hrs for CTAS 3… • Heavy clinical load – Do the authors speculate that this will be explicitly different from understaffing and overcrowding or

	confounded by the latter two? Who will determine presence of "heavy clinical load"? I would be more specific: If a TTL or Major MD or Team Lead RN feels that they are " overwhelmed"… or if any of them on a scale 0-10 (light load) 5 (average level), 10 (completely overwhelmed), select 7 or more…we would abort the in-situ simulation. • Low bed availability on wards – Who will decide on evoking a "no go" with that regard? What magnitude of bed deficit constitutes a "low availability"? Will you only know about it by experiencing overcrowding in your department? Or increased number of admitted patients in your rooms and hallways? Why otherwise low number of available beds on the wards would affect the ED and prevent you from carrying out an in-situ simulation? I would remove this as a "no go" criterion. • Equipment needs (e.g., unavailability of the fast flow fluid warmer). • If a real trauma activation is expected or ongoing • Unanticipated Events/Threats to Psychological Safety – I would add "…and Physical Safety" (I presume that you mean fire, flood, violent threats etc.) The authors may choose not to address my comments in the revision of the proposal. I do, however, believe that if they omit having a detailed description of the reasons to abort the in-situ simulation in the final paper, it may be viewed as a potential weakness of the study and a threat to its external validity.

VERSION 3 – AUTHOR RESPONSE

Reviewer 1

I am grateful for the author's responses to my questions, their clarifications and revisions. I am satisfied with their responses and now have a much clearer understanding of the project. As the general assumption goes, most latent and overt safety threats will occur exactly during the times when system is stretched to the limit. Unfortunately, those will be exactly the times when the authors propose to abort in-situ simulations. I do appreciate that ultimately, patients' and clinicians' safety take priority over research and training. Thus, my only remaining suggestion is to clarify or to be more descriptive of what constitutes "No go criteria" that will lead to cancellation of the in -situ simulations? I am following the list provided by the authors in the appendix: • Availability of the environment – I am not sure what the authors mean by that? • Medical Understaffing –How many RNs or MDs or RTs in your department will it take to be absent to declare "understaffing"? I would specify the exact number: e.g. if the department is more than	Thank you for your comments. As per your suggestion, we have added further details to our list of ''no go'' criteria (see Online supplementary material).

1 RN short or at least 1 MD did not show up, or an on-call MD needed to be called…etc. • Non-medical Understaffing – Do you mean environmental support workers or administrative staff or scribes? How many of them will it take to declare “understaffing”? Who will decide whether the department I understaffed by non-clinical personnel? Why would it affect a clinical in-situ simulation? Unless you have a strong opinion about that, I would remove that criterion. • Overcrowding - Do the authors have a predetermined wait time in mind? Or the number of people in the waiting room? Or the number of admitted patients in the department? I would specify: e.g. wait time > 2 hrs for CTAS 3… • Heavy clinical load – Do the authors speculate that this will be explicitly different from understaffing and overcrowding or confounded by the latter two? Who will determine presence of “heavy clinical load”? I would be more specific: If a TTL or Major MD or Team Lead RN feels that they are “ overwhelmed”… or if any of them on a scale 0-10 (light load) 5 (average level), 10 (completely overwhelmed), select 7 or more…we would abort the in-situ simulation. • Low bed availability on wards – Who will decide on evoking a “no go” with that regard? What magnitude of bed deficit constitutes a “low availability”? Will you only know about it by experiencing overcrowding in your department? Or increased number of admitted patients in your rooms and hallways? Why otherwise low number of available beds on the wards would affect the ED and prevent you from carrying out an in-situ simulation? I would remove this as a “no go” criterion. • Equipment needs (e.g., unavailability of the fast flow fluid warmer). • If a real trauma activation is expected or ongoing • Unanticipated Events/Threats to Psychological Safety – I would add “…and Physical Safety” (I presume that you mean fire, flood, violent threats etc.) The authors may choose not to address my comments in the revision of the proposal. I do, however, believe that if they omit having a detailed description of the reasons to abort the in-situ simulation in the final paper, it may be viewed as a potential weakness of the study and a threat to its external validity.	